

A BIPOLAR JUNCTION TRANSISTOR HAVING A HIGH GERMANIUM
CONCENTRATION IN A SILICON-GERMANIUM LAYER AND A METHOD
FOR FORMING THE BIPOLAR JUNCTION TRANSISTOR

[0001] This patent application claims the benefit of United States Provisional Patent Application No. 60/552,308 filed on March 10, 2004.

FIELD OF THE INVENTION

[0002] The present invention relates to silicon-germanium bipolar junction transistors and more specifically, to a method for fabricating a silicon-germanium transistor comprising a germanium-enriched region and a bipolar junction transistor comprising a germanium-enriched region.

BACKGROUND OF THE INVENTION

[0003] Bipolar junction transistors are commonly employed in integrated circuits that require high-speed amplifiers or high-speed switches. A bipolar junction transistor (BJT) comprises three adjacent doped semiconductor regions having an NPN or PNP doping configuration. A middle region forms a base and two end regions separated by the base form an emitter and a collector. The middle base region is physically narrow relative to the minority carrier diffusion length for carriers within the base. Typically, the emitter has a higher dopant concentration than the base and the collector, and the base has a higher dopant concentration than the collector. A small signal applied to one of the BJT terminals modulates large changes in current through the other two terminals. A BJT operates to amplify an input signal supplied between the base and the emitter, with the output signal appearing across the emitter/collector. The BJT can also operate as a switch with an input signal applied across the base/emitter junction switching the emitter/collector circuit to an opened or a closed (i.e., short-circuited) state.

[0004] The emitter current primarily comprises the injection of carriers from the emitter into the base, which is achieved by making the donor concentration in the emitter much greater than the acceptor concentration in the base. Thus for the

common NPN BJT, electrons are injected into the base with negligible hole injection into the emitter from the base. Since the base is very narrow compared to the minority carrier diffusion length (the diffusion length of the electrons in the base), the carriers injected into the base do not recombine in the base, but diffuse across the base into the reverse-biased base-collector junction. Thus a current across the reverse-biased base-collector junction is determined by the carriers injected from the emitter that arrive at the base-collector depletion region. The dopant concentration in the collector is less than that in the base, so the depletion region extends primarily into the collector.

[0005] There are several known semiconductor fabrication processes for forming the three doped regions of a bipolar junction transistor, and several different BJT architectures can be formed according to these processes. The simplest structure comprises a planar architecture having stacked NPN or PNP regions formed by successive dopant implants into a silicon substrate.

[0006] Significant performance enhancements are achieved by a heterojunction bipolar junction transistor (HBT) having a silicon-germanium base. It is known that the silicon-germanium base exhibits a narrower band gap and lower resistivity than a silicon base. Thus the HBT provides improved high-speed and high-frequency operation over the conventional BJT. Increasing the germanium concentration in the silicon-germanium base results in a larger valence band offset between the emitter and the base, leading to enhanced bulk electron and hole mobility, further improving high-speed/high-frequency operation. At a germanium concentration of about 20%, the valence band offset is about 0.17 eV.

[0007] Prior art methods for forming an epitaxially grown layer of silicon-germanium overlying a silicon layer (e.g., a silicon-germanium base overlying a silicon collector) carefully control a temperature, a pressure and a reactive gas flow rate during epitaxial growth to achieve germanium concentrations of about 10% to 25% (i.e., about 90% to 75% silicon) in the silicon-germanium layer. As the germanium concentration increases, compressive strain in the silicon-germanium layer increases. Crystalline dislocations form to relieve the strain. The number of dislocations increases as the germanium concentration increases, eventually reaching a level where the dislocations disrupt the epitaxial properties of the silicon-germanium layer, negating the advantageous properties of the silicon-germanium layer. Thus the germanium concentration must be limited to limit the number of dislocations.

[0008] Use of a buffer layer (wherein the germanium concentration is varied gradually, with the germanium concentration increasing in a direction away from the collector) between the silicon collector and the silicon-germanium base somewhat reduces strain relaxation and may thereby aid in achieving these concentration levels.

[0009] It is known that crystalline defects in a transistor can limit performance. In particular, base region defects, such as the dislocations described above, can reduce the transistor cut-off frequency, current gain and maximum oscillating frequency.

SUMMARY OF THE INVENTION

[0010] According to one embodiment the present invention comprises a method of manufacturing a semiconductor device further comprising: epitaxially growing a silicon-germanium base on a collector, thermally oxidizing the base to preferentially grow silicon dioxide on an upper surface of the base to form a germanium-enriched region in an upper region of the base, removing the silicon dioxide and depositing an emitter overlying the base.

[0011] Another embodiment of the present invention comprises a heterojunction bipolar transistor comprising: a collector; a base disposed above the collector, the base comprising a silicon-germanium layer; a germanium-enriched region proximate an upper surface of the base and within the silicon-germanium layer and an emitter overlying the germanium-enriched region.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The foregoing and other features of the invention will be apparent from the following more particular description of the invention, as illustrated in the accompanying drawings, in which like reference characters refer to the same parts throughout the different figures. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

[0013] Figures 1 - 7 are cross-sectional illustrations through a common plane illustrating sequential processing steps for forming a silicon-germanium bipolar transistor according to the teachings of the present invention.

[0014] Figure 8 qualitatively illustrates a base dopant profile for a silicon-germanium transistors according to the present invention.

[0015] Figures 9 - 11 are cross-sectional illustrations through a common plane illustrating additional sequential processing steps for forming a silicon-germanium bipolar transistor according to the teachings of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0016] Before describing in detail a method for forming a heterojunction bipolar transistor and a heterojunction bipolar transistor structure, it should be observed that the present invention resides primarily in a novel and non-obvious combination of elements and process steps. Accordingly, the inventive elements and steps have been represented by conventional elements and steps in the drawings, showing only those specific details that are pertinent to the present invention so as not to obscure the disclosure with details that will be readily apparent to those skilled in the art having the benefit of the description herein.

[0017] The present invention teaches a method for achieving a higher germanium concentration than attainable according to the prior art methods, in the base region of a silicon-germanium bipolar transistor. According to one embodiment, thermal oxidation of a silicon-germanium base (or silicon-germanium portion of the base) forms an upper silicon-dioxide layer and a substantially defect-free (i.e., substantially free of dislocations) germanium-enriched silicon-germanium region proximate the silicon dioxide/silicon-germanium interface. The germanium-enriched region has a greater germanium concentration (on the order of 30% to 75% germanium) than the germanium concentration outside the enriched region and a greater concentration than the in an HBT base of the prior art.

[0018] The higher germanium concentration within the enriched region is achieved without substantial strain relaxation (i.e., without the formation of a significant number of dislocations) and produces a greater valence band offset that imparts advantageous operating characteristics to the transistor. The method of the present invention may also eliminate the need for a buffer layer between the silicon and silicon-germanium layers and thus eliminates fabrication steps.

[0019] A germanium-enriched region of the present invention with a concentration of about 50% germanium provides a valence band offset of about 0.37 eV and a germanium concentration of about 30% germanium provides a valence band offset of about 0.21 eV, both providing commensurate operating speed increases compared to

lower valence band offsets in the prior art. Increasing the germanium base concentration from about 20% to about 50% increases the transistor current gain (β) by about three orders of magnitude.

[0020] The teachings of the invention can be applied to a silicon-germanium base having a uniform dopant (i.e., germanium) profile, a stepped dopant profile or any graded dopant profile, including a graded dopant profile having a higher germanium concentration proximate the collector relative to a concentration in a portion of the base proximate the emitter. The teachings of the present invention can also be applied to other base dopant profiles, including a layered base having two or more different dopant profiles (in one embodiment discontinuous dopant profiles) within the base layers. One example of such a layered base dopant profile comprises a first dopant profile in a lower one-third layer of the base overlying the collector (wherein the dopant concentration decreases in a direction away from the collector), a second dopant profile within a middle one-third layer of the base and a third dopant profile in an upper one-third layer of the base. Use of such non-uniform base dopant profile in conjunction with the present invention may facilitate creating an optimum germanium concentration in the germanium-enriched region. Further, the dopant profile and the concentrations of the silicon-germanium base can be selected such that the germanium-enriched region formed within the base optimizes desired transistor-operating parameters according to an intended application.

[0021] A process sequence for forming a germanium-enriched region according to the teachings of the present invention for an NPN HBT is described below in conjunction with the Figures 1-7 and 9-11, which show formed structures through the same cross-sectional plane.

[0022] The illustrated process forms an NPN HBT on a substrate 10 of Figure 1. To avoid performance degradation and electrical cross-talk between integrated circuit devices, it is necessary to electrically isolate the HBT from other circuit devices. The illustrated exemplary isolation process comprises LOCOS (local oxidation of silicon) isolation regions 20. It is known that in other embodiments, other isolation structures, such as shallow and deep trench isolation structures, can be used alone or in combination with the silicon dioxide isolation regions 20. A silicon dioxide layer 26 is formed over the substrate 10 between the isolation regions 20.

[0023] Figure 1 also illustrates a subcollector 30 formed within the substrate 10 by a dopant implant process through an appropriately patterned implant mask. A subsequent masking and implant operation forms an n-type lightly doped collector contact region 40.

[0024] In Figure 2, a TEOS silicon dioxide spacer layer 50 and a polysilicon layer 55 are formed on the substrate 10 according to known processes. The polysilicon layer 55 is doped with a high-dose boron implant (as represented by implant arrowheads 65) through an implant mask 60. After a later patterning step, the boron-implanted polysilicon layer 55 forms an extrinsic base as described below.

[0025] As illustrated in Figure 3, a silicon nitride layer 70 and a silicon dioxide layer 75 (in one embodiment formed according to a TEOS process) are deposited over the polysilicon layer 55. A photoresist layer 90 is deposited and patterned to form a window 100 therein. The silicon dioxide layer 75, the silicon nitride layer 70 and the polysilicon layer 55 are anisotropically etched through the window 100, stopping on the TEOS layer 50, to form an emitter window 110 in the substrate 10. A collector region 120 is implanted through the window 100.

[0026] After removing the photoresist layer 90 and the silicon dioxide layer 75, a layer of silicon nitride is deposited and anisotropically etched to form sidewall spacers 170. See Figure 4. A wet etch process removes the silicon dioxide TEOS layer 50 and the silicon dioxide layer 26 from within the emitter window 110, forming a primary cavity 174 and cavities 175 laterally disposed relative to the primary cavity 174.

[0027] As illustrated in Figure 5, a silicon-germanium base 180 is formed in the cavities 174 and 175 (see Figure 4) during a silicon-germanium epitaxial growth step. The silicon-germanium base 180 is typically grown according to a chemical vapor deposition (CVD) reactor process to provide a desired proportion of germanium-to-silicon in the base. According to different embodiments of the invention, the silicon-germanium base 180 comprises a graded germanium dopant profile, a uniform germanium dopant profile or a stepped germanium dopant profile.

[0028] As illustrated in Figures 6 and 7, the silicon-germanium base 180 is thermally oxidized to form a thermal oxide layer 185 (e.g. silicon dioxide) over the silicon-germanium base 180. Dry or wet oxidation can be employed to form the thermal oxide layer 185. Preferably, the dry oxidation is performed within a temperature range from about 700 °C to about 900 °C at atmospheric pressure and an oxygen flow rate of

approximately 2 liters/min. According to other embodiments, pressures below atmospheric can be used. An exemplary oxidation process comprises the use of dry oxidation at about 900 °C for about 1 hour at atmospheric pressure with an oxygen flow rate of about 2 liters/min. Conventional wet oxidation processes can be employed in lieu of dry oxidation.

[0029] During oxidation, a low-defect density germanium-enriched region 200 (having a thickness of about 3 to 5 nanometers) is formed below an upper surface 190 of the silicon-germanium base 180. Various oxidation process parameters (e.g., duration, temperature and pressure) can be varied to alter the thickness and germanium concentration of the germanium-enriched region 200. In an exemplary embodiment of the present invention, a germanium enriched region 200 approximately three to four nanometers thick is formed in the silicon-germanium base 180 with a thickness of approximately 120 nanometers. As illustrated in Figure 7, the germanium-enriched region 200 ranges from a region 225 encompassing an area of the silicon-germanium base 180 exposed to the thermal oxide layer 185, to a region 250 including the entire silicon-germanium base 180.

[0030] During thermal oxidation of the silicon-germanium base 180, the thermal oxide layer 185, typically silicon dioxide, but not necessarily stoichiometric, is preferentially produced relative to germanium dioxide, resulting in formation of more silicon dioxide than germanium dioxide. The resulting thermal oxide layer 185 comprises primarily silicon dioxide. During oxidation, germanium in the oxidizing region diffuses or otherwise migrates in a direction of the thermal oxide layer 185 to form the germanium-enriched region 200. It is believed that unbound germanium accumulates into a reformed lattice within the germanium-enriched region 200 (also referred to as a thermally oxidized germanium-enriched region) creating a relatively high germanium concentration of as much as five or more times the original germanium concentration in the as-grown silicon-germanium base 180. Further, the germanium-enriched region 200 comprises a low-defect density single crystal lattice including desirable compressive strain properties. The germanium-enriched region 200 has crystalline properties substantially similar to an epitaxially grown layer and may be substantially defect-free, despite the presence of a high germanium concentration. Thus the present invention employs an oxidation process during HBT processing to achieve a high germanium concentration in an epitaxially grown silicon-germanium base.

[0031] Figure 8 qualitatively illustrates an approximate concentration profile 216 for one embodiment of the germanium-enriched region 200, noting the abruptness of the concentration reduction in a direction away from the emitter. The dopant profile 216 is preferably in a range of 30% to 75% germanium concentration or higher. In the illustrated embodiment, the germanium dopant profile outside the germanium-enriched region 200 is similar to the germanium dopant profile in the base 180 prior to the thermal oxidation step.

[0032] After oxidation of the silicon-germanium base 180 and formation of the germanium-enriched region 200, the thermal oxide 185 is removed or stripped using standard semiconductor processing techniques such as a hydrofluoric acid etch. A resultant structure is illustrated in Figure 9.

[0033] Silicon nitride spacers 260 and underlying silicon dioxide (TEOS) spacers 270 are formed in the window 110 as illustrated in Figure 10. The spacers, which serve to increase a distance between a later-formed emitter (having an n⁺ doping in an NPN transistor) and an extrinsic base (having a p⁺ doping in an NPN transistor) are formed by depositing a TEOS silicon dioxide layer and an overlying silicon nitride layer. The layers are anisotropically etched back to form the spacers 260 and 270 as illustrated, with the etch stopping on a region of the TEOS silicon dioxide layer formed on an upper surface of the base 180. In another embodiment the spacers 260 and 270 may not be required if the previously formed spacers 170 provide sufficient isolation. Following spacer formation, the remaining region of the TEOS silicon dioxide layer overlying the upper surface of the base 180 is removed by a wet etch process.

[0034] An emitter layer 280 is grown or deposited on the base 180 as shown in Figure 10 and doped by implant or in-situ doping. Typically, the emitter layer 280 comprises a crystalline or polycrystalline emitter layer that is epitaxially grown or otherwise deposited according to techniques known in the art. A hard mask and photoresist mask are utilized to pattern the doped emitter layer 280 to form an emitter 280A as illustrated in Figure 11. Regions of the silicon nitride layer 70 are also removed during the emitter region etch, leaving silicon nitride regions 70A underlying the emitter 280A. Using another photoresist mask, the TEOS layer 50 and the polysilicon layer 55 are etched, with the latter forming an extrinsic base 55A.

[0035] A final HBT 300 comprises a substantially defect-free germanium-enriched region 200 having a germanium concentration greater than is achievable by prior art standard epitaxial growth methods.

[0036] According to another embodiment of the present invention, the HBT can be annealed to redistribute the germanium atoms in the germanium-enriched region 200, lowering the germanium concentration in the enriched region 200 and raising the germanium concentration in a region of the base proximate the germanium-enriched region 200.

[0037] The method of the present invention can be applied to the fabrication of various microelectronic devices that can benefit from a low-defect level, high-germanium concentration in a silicon-germanium region.

[0038] A feature of a disclosed embodiment of the invention is an HBT having a germanium-enriched region in the HBT base. Another feature comprises growing a thermal oxide layer over a silicon-germanium base to enrich the germanium concentration of a region of the base and removing the thermal oxide layer.

[0039] The present invention, due to the relatively high germanium concentration in the germanium-enriched region, improves the circuit designer's ability to optimize HBT operating parameters (e.g., current gain, cutoff frequency, maximum oscillation frequency and gate delay) for a specific design application. For example, a higher germanium fraction (i.e., a ratio of the germanium concentration to the silicon concentration) in the HBT base raises the current gain and the cutoff frequency.

[0040] Although described in the context of an integrated circuit having BJTS formed therein, the teachings of the present invention can also be applied to a process for forming BJTS/HBTS in a BiCMOS process, wherein BJTS/HBTS and complimentary metal oxide field effect transistors are formed in a substrate.

[0041] An HBT architecture comprising a germanium-enriched region in a silicon-germanium base and a process for forming a germanium-enriched region in a silicon-germanium base in an HBT have been described. Specific applications and exemplary embodiments of the invention have been illustrated and discussed, which provide a basis for practicing the invention in a variety of ways and in a variety of circuit structures. Numerous variations are possible within the scope of the invention. Features and elements associated with one or more of the described embodiments are not to be

construed as required elements for all embodiments. The invention is limited only by the claims that follow.